AD_____

Award Number: W81XWH-11-1-0342

TITLE: Identification of Genes and Genetic Variants Associated with Poor Treatment Response in Patients with Prostate Cancer

PRINCIPAL INVESTIGATOR: Lisa A. Cannon-Albright

CONTRACTING ORGANIZATION: University of Utah
Salt Lake City, UT 84108-1266

REPORT DATE: October 2013

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
                        Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
                        Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2013	Annual	30September2012-29September2013

4. TITLE AND SUBTITLE

Identification of Genes and Genetic Variants Associated with Poor Treatment Response in Patients with Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-1-0342

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Lisa A. Cannon-Albright, Neeraj Agarwal, Craig Teerlink

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

E-Mail: lisa.albright@utah.edu

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Utah

Department of Internal Medicine
391 Chipeta Way Suite D
Salt Lake City, UT 84108-1266

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Our initial linkage analysis identified a significant linkage on chromosome 11 and a manuscript is in preparation. We will now collaborate (with other funding) to begin sequence analysis of the regions of interest. We have expanded our identification of high-risk pedigrees and identified an additional 192 samples in new high risk pedigrees soon to be genotyped. Some of these samples extend previously identified pedigrees with linkage evidence. We have begun sequence analysis of 34 lethal prostate cancer cases and identified a set of candidate genes in a set of 6 pairs of lethal prostate cancer cases sequenced.

Identification of genes predisposing to recurrent/lethal prostate cancer from this study will validate this powerful approach, which can be extended to other high risk prostate cancer pedigrees, and will identify genes and pathways that can be further examined to expand our knowledge of prostate cancer genetics.

15. SUBJECT TERMS
Recurrent/lethal prostate cancer, linkage analysis, genes, pathways, prostate cancer genetics, chromosome 1, chromosome 11

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			
U	U	U	UU	21	19b. TELEPHONE NUMBER (include area code)

Table of Contents

Introduction

Decades of investigation into the genetic causes of prostate cancer (prca) and prostate cancer aggressiveness has yet to clearly identify genes or variants which explain much more than a small amount of risk for prostate cancer among a small population of men. Even less progress has been made in understanding why 30% of all patients with localized prostate cancer eventually develop recurrent, and subsequently fatal, prostate cancer, or in understanding the factors that are associated with the range of treatment response (survival time) after diagnosis and treatment for recurrent prostate cancer.

Here we propose a genetic study of the distinct subset of recurrent prostate cancer cases: those who will, in all likelihood, go on to die from their prostate cancer. We will further stratify and study these cases by their response to castration (chemical or physical) treatment, the standard of care for patients with recurrent prostate cancer. Using a unique and powerful statewide, population-based resource, we will identify and sample over 800 recurrent cases of prostate cancer. We will perform genome wide genotyping on informative cases in high-risk pedigrees, and we will apply complementary genetic analyses to identify genes and variants predisposing to recurrent prostate cancer, and variable response to treatment.

Body

Task 1. Recruitment and sampling of new and returning recurrent prostate cancer patients at HCI

We have recruited and sampled 174 recurrent prostate cancer cases attending Dr. Agarwal's clinic at HCI. A DNA sample and prostate cancer questionnaire has been collected and stored for each. The prostate cancer characteristics phenotype form is in the Appendix.

We have also identified a total of 60 high-risk prostate cancer kindreds with an excess of lethal prostate cancer and have identified an additional 192 DNA samples for these historical cases in pedigrees that we are preparing for genotyping now. (192 samples were genotyped already in Year 1)

> **sub task 1a. Assignment of phenotype for treatment response for recruited cases**
>
> Dr. Agarwal's fellow has begun to review charts and enter phenotypes (see the Appendix) for each case recruited in clinic. This year we will genotype these cases and perform GWAS for prostate cancer characteristics and treatment response.

Task 2. Identification of most informative samples for genotyping

We have identified 60 high-risk informative lethal prostate cancer pedigrees with at least 3 samples available for genotyping. We genotyped 192 (96 x 2) prostate cancer cases in these high-risk informative recurrent prostate cancer pedigrees last year, and are now preparing the next set of 192 for genotyping. These represent prostate cancer cases who died with prostate cancer contributing to their death who are members of high-risk prostate cancer pedigrees with a at least 3 lethal prostate cancer cases already sampled.

Task 3. Genotype 200 samples each year

We are just preparing the 192 (2 plates x 96 samples/plate) samples for genome wide genotyping for this year from the high-risk lethal/recurrent prostate cancer pedigrees discussed for Task 2. We will also genotype the 174 samples collected in Dr. Agarwal's clinic. (Since Illumina has just cut prices by ½, we will be able to genotype these additional cases). We will perform genome wide genotyping at the University of Utah Genotyping Core Facility using the Illumina 720,000 SNP Omni Express set of markers.

Task 4. Import and quality control of genotype data

We acquired genotype data for the initial 196 samples from last year. All data has undergone standard quality control and was imported into our family study database and linked to phenotype data.

Task 5. Association analysis of all genotyped samples

> **subtask 5.1   Selection of genotypically matched controls**
>
> Using existing software we have analyzed available iControl public data from Illumina for all Caucasians and selected an appropriate genotypically matched set of controls for Utah prostate cancer cases (see Methods in Teerlink et al., 2011).
>
> **subtask 5.2   Association Analysis**
>
> Because we have not completed genotyping of the recurrent prostate cancer cases seen in Dr. Agarwal's clinic, and because association analysis requires large sample sizes, we have not begun to analyze the data for associations. We will use existing public software that allows correction for relatedness of cases (GEMMA) when we have additional genotype data this year for the 174 recurrent prca cases seen in clinic and the 192 cases in high-risk pedigrees.

subtask 5.3   Validation

We will collaborate with the International Consortium for Prostate Cancer Genetics (ICPCG) and other consortia who are also performing association analysis for aggressive prostate cancer to perform a validation of our findings after analysis in year 2. We have two collaborators in the ICPCG who have also begun to focus on lethal prca as a subset of interest and who have begun sequencing and genotyping.

Task 6. Linkage analysis of all genotyped samples from informative pedigrees

subtask 6.1   Selection of markers for linkage

A set of 301, 646 markers with no linkage disequilibrium was selected for linkage analysis. From the intersection of SNP markers from the five Illumina genotyping platforms that we have used (550K, 610K, 1M,Omni_express (720k), and Omni_1M).   We selected a set of 25,436 SNP genome wide markers for linkage with good chromosomal representation and low/no LD (Cannon-Albright et al., 2012).
This same set of markers will be used for the linkage analysis of the new 192 linkage samples being readied for genotyping now.

subtask 6.2   Linkage analysis

We performed genome-wide linkage analysis for 21 of the 27 already genotyped high risk lethal prostate cancer pedigrees in year 1. We are adding new lethal high risk pedigrees to our linkage set and additional cases have been added to informative pedigrees to explore these findings.

One pedigree has already provided significant evidence for linkage (LOD > 3.30) for the dominant mode of inheritance. This pedigree has 4 lethal prostate cancer cases and 10 additional prostate cancer cases genotyped. The segregating chromosome 11 haplotype providing linkage evidence is shown in Figure 2.

We have identified 3 additional lethal prca cases to genotype in this pedigree and they will be genotyped in the new set of 192 cases we are preparing for genotyping now.

We observed evidence for dominant linkage at a chromosome 1 locus (LOD > 3.0) for which we have previously published linkage evidence (Camp et al., 2005). We sequenced the initial pedigree with significant linkage reported previously in a collaboration with ICPCG. We have received the sequence data for the 3 cases in the chromosome 1 linked pedigree. We have identified one more sampled lethal prca case in this pedigree which will be genotyped and used for confirmation of sequence variants identified.
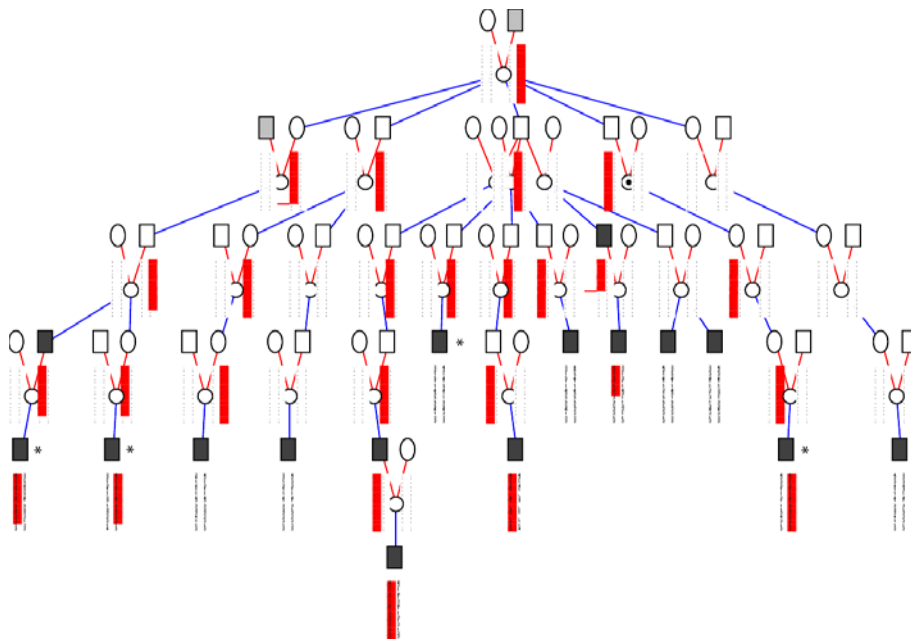
Figure 2. **Prostate cancer pedigree with significant evidence for linkage to chromosome 11 (LOD = +3.56). Lethal prostate cancer cases are marked with *. Affected males are fully shaded. The red (dark) haplotype on chromosome 11 that segregates with prostate cancer is shown.**

Sequence Analysis

We have completed genome wide exomic sequencing for 2 lethal prca cases in this pedigree and bioinformatic sequence analysis is underway in the specific region with evidence for linkage.

We have completed genome wide exomic sequencing of an additional 5 pairs of lethal prostate cancers who are cousins and who each have an affected first degree relative. Bioinformatic analysis is now underway.

Finally in a collaboration with International Consortium for Prostate Cancer Genetics (ICPCG), we have also had exomic sequencing performed on 121 prostate cancer cases in high risk pedigrees. Twenty two of these cases are also "lethal" prostate cancer cases whose cause of death was prostate cancer. We have just received summary files from the ICPCG and will also perform bioinformatics analysis of these lethal prostate cancer cases in search of prostate cancer predisposition genes or variants.

Bioinformatic Analysis of Sequence Variation for 12 lethal prostate cancer cases.

Each sequencing run produces an enormous number of sequence reads that must be aligned to the human genome. Aligned reads are then analyzed to determine the location of genetic variants exist in each sample. Variants are then prioritized for their potential impact on the disease. Raw sequence data was generated at Huntsman Cancer Institute on the Illumina hi-seq2 high-throughput sequence analyzer with the Illumina TruSeq target library. Individual sequence reads coming from the sequence instrument were mapped back to the human reference genome (hg 19) and Novoaligner software which uses uses pairing information from paired end reads to more accurately assign reads with multiple matches to the genome. We used The Broad Institutes's Genome Analysis Tool Kit (GATK) for variant calling of single nucleotide variants (SNVs) and small insertions or delestions (indels). We followed the Broad Institute's best practice guidelines concerning re-aligment and recalibration of aligned reads in order to produce the most reliable set of called variants. GATK software also provides a rich set of quantitative variant annotations, which it uses to estimate the probability that the called variants are true positives, thus providing an optimally derived set of SNVs and short indels. We used the Annovar software package to further annotate called variants.

After a reliable set of possible variants across all samples was derived from the above process, we used a set of filtering rules to derive a set of potential candidates for lethal prostate cancer susceptibility. Our filtering strategy was as follows: we eliminated variants with read depth less than 10 reads; that are not shared by both members of a sequenced pair; that have a high observed frequency ($> 1\%$) in 1000 Genomes project or the NHLBI Whole Exome project. We then prioritized variants according to their predicted post translational impact. Tier 1 variants are those classified as frameshift mutations (insertions or deletions) and non-sense mutations (stop-gain SNVs or stoploss SNVs). Tier 2 variants include missense mutations that were characterized as damaging (score > 0.99) by the MutationTaster software package (Schwarz 2010). MutationTaster incorporates information from several other variant scoring freely available software packages to determine the expected potency of the mutation. All other variants passing the filtering scheme will be assigned to Tier 3, and will be considered the lowest priority variants for further investigation.

As a further refinement to our variant filtration scheme, we used evidence that a variant was inherited from a common ancestor, derived from shared genomics segment (SGS) analysis. SGS analysis, developed by the PI, essentially counts the number of contiguous SNPs that could be shared between a set (in this case a pair) of related cases of interest (Thomas et al., 2008). Long runs of SNPs indicate regions likely to be inherited from a common ancestor (identical-by-descent, or IBD). The distribution of identical-by-state (IBS) sharing also emerges from this analysis. Hence, IBD regions of the genome can be distinguished from IBS regions via SGS analysis. Simulation techniques can be employed to assess significance of findings. When sequenced cases are from the same high-risk pedigree, we can reduce our search for variants to those regions that appear IBD among cases. In this analysis, variants between a related pair (all were approximately cousins) of sequenced lethal prostate cases that occurred in a genetic segment denoted as likely to be IBD were given highest priority. Evidence for IBD sharing is consistent with our assumption that the variants we are attempting to identify convey predisposition and

should occur in multiple affected people in the pedigree. We used high-density SNP genotype data to conduct SGS analysis in the 12 lethal prostate cancer cases.

A brief summary of the Tier 1 mutations identified in the 12 sequenced lethal prostate cases that meet all of these criteria appear in Table 1. We did not see any frameshift mutations that were shared by both pedigree members in any pedigree, but we did detect 4 stop-gain mutations that met this criteria. Only the variant in GPATCH 2 appears in a genomic region identified as IBD and is our candidate of highest interest at this point of our analysis. This, and other Tier 1 and Tier 2 variants, will next be confirmed by Sanger sequencing.

Table 1. Stop-gain mutations passing the variant filtering scheme post bioinformatics analysis.

Pedigree	Band	bp position (b37)	dbSNP name	MAF in 1K Genomes	SGS segment length	Gene
3	1q41	217,604,657	NA	NA	2805	GPATCH2
5	1p36	17,034,125	rs141324796	0	26	ESPNP
6	2q34	209,302,328	rs617423329	A=0.0082	110	PTH2R
6	10q21	61,122,268	rs3078330	0	27	FAM13C

subtask 6.3   Validation

Most prostate cancer linkage studies do not use the recurrent/lethal phenotype that we use, so these findings may be difficult to validate. We will continue to review all prostate cancer linkage reports and contact appropriate groups to attempt to validate our regions of interest in high risk prostate cancer pedigrees.
Two groups of collaborators in the ICPCG have told us that they are similarly focused on the subset of lethal prostate cancer cases and are moving forward with genotyping and sequencing; one of these groups has a significant GWAS hint at our chromosome 11p region in lethal prca cases.

Task 7. Publication of linkage and association manuscripts

A publication reporting the significant linkage finding on chromosome 11 is in preparation; it has proven difficult to publish linkage evidence on its own and we are hoping to supplement with some sequence analysis. We have published analysis of the UPDB showing that there is strong evidence that the subset of lethal prostate cancer cases cluster more in pedigrees than all prostate cancer cases. These results form the basis of our hypothesis that analysis of these homogeneous pedigrees will result in predisposition gene identification (Nelson et al., 2013).

Key Research Accomplishments

- Creation of a set of 60 high risk prostate cancer pedigrees with DNA samples representing an excess of the most clinically significant subset of prostate cases: those with recurrent/lethal disease

- Initial linkage analysis of less than 200 cases has already identified a new significant linkage on chromosome 11; new pedigrees have been identified and extensions to initial pedigrees. Genotyping of this new set of 192 cases will begin next month

- We have sampled 174 newly recurrent prostate cancer cases and collected detailed cancer characteristics data.

Reportable outcomes

- initial linkage manuscript is in preparation-may add sequence data
- sequence analysis of exome data for 34 lethal cases is underway

Conclusions

The first 2 years of this grant has already resulted in an informative set of DNA, high risk pedigrees, and phenotype data for a set of pedigrees representing an excess of a highly significant clinic subset of prostate cancer cases: those who will go on to die of the disease.

We have already identified significant evidence for linkage and have found collaborators and begun sequence analysis of the regions of interest and 34 lethal prostate cancer cases.

Identification of genes predisposing to recurrent/lethal prostate cancer from this study will validate this powerful approach, which can be extended to other high-risk prostate cancer pedigrees, and will identify genes and pathways that can be further examined to expand our knowledge of prostate cancer genetics.

References

Amin Al Olama A, Kote-Jarai Z, Schumacher FR, Wiklund F, Berndt SI, Benlloch S, Giles GG, Severi G, Neal DE, Hamdy FC, Donovan JL, Hunter DJ, Henderson BE, Thun MJ, Gaziano M, Giovannucci EL, Siddiq A, Travis RC, Cox DG, Canzian F, Riboli E, Key TJ, Andriole G, Albanes D, Hayes RB, Schleutker J, Auvinen A, Tammela TL, Weischer M, Stanford JL, Ostrander EA, Cybulski C, Lubinski J, Thibodeau SN, Schaid DJ, Sorensen KD, Batra J, Clements JA, Chambers S, Aitken J, Gardiner RA, Maier C, Vogel W, Dork T, Brenner H, Habuchi T, Ingles S, John EM, Dickinson JL, **Cannon-Albright L**, Teixeira MR, Kaneva R, Zhang HW, Lu YJ, Park JY, Cooney KA, Muir KR, Leongamornlert DA, Saunders E, Tymrakiewicz M, Mahmud N, Guy M, Govindasami K, O'Brien LT, Wilkinson RA, Hall AL, Sawyer EJ, Dadaev T, Morrison J, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As N, Woodhouse CJ, Thompson A, Dudderidge T, Ogden C, Cooper CS, Lophatonanon A, Southey MC, Hopper JL, English D, Virtamo J, Le Marchand L, Campa D, Kaaks R, Lindstrom S, Diver WR, Gapstur S, Yeager M, Cox A, Stern MC, Corral R, Aly M, Isaacs W, Adolfsson J, Xu J, Zheng SL, Wahlfors T, Taari K, Kujala P, Klarskov P, Nordestgaard BG, Roder MA, Frikke-Schmidt R, Bojesen SE, FitzGerald LM, Kolb S, Kwon EM, Karyadi DM, Orntoft TF, Borre M, Rinckleb A, Luedeke M, Herkommer K, Meyer A, Serth J, Marthick JR, Patterson B, Wokolorczyk D, Spurdle A, Lose F, McDonnell SK, Joshi AD, Shahabi A, Pinto P, Santos J, Ray A, Sellers TA, Lin HY, Stephenson RA, Teerlink C, Muller H, Rothenbacher D, Tsuchiya N, Narita S, Cao GW, Slavov C, Mitev V, Chanock S, Gronberg H, Haiman CA, Kraft P, Easton DF, Eeles RA. (2013). A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. *Hum Mol Genet*, *22*(2), 408-15.

Camp NJ, Farnham JM, **Cannon-Albright LA**. (2005). Genomic search for prostate cancer predisposition loci in Utah pedigrees. *Prostate*, *65*(4), 365-74.

Cannon-Albright LA, Teerlink CC, Farnham JM, Thomas AW, Zone JJ, Leachman SA. Linkage analysis of extended high-risk pedigrees replicates a cutaneous malignant melnaoma predisposition locus on chromosome 9q21. J Invest Dermatol Sept 6 2012.

Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Ghoussaini M, Luccarini C, Dennis J, Jugurnauth-Little S, Dadaev T, Neal DE, Hamdy FC, Donovan JL, Muir K, Giles GG, Severi G, Wiklund F, Gronberg H, Haiman CA, Schumacher F, Henderson BE, Le Marchand L, Lindstrom S, Kraft P, Hunter DJ, Gapstur S, Chanock SJ, Berndt SI, Albanes D, Andriole G, Schleutker J, Weischer M, Canzian F, Riboli E, Key TJ, Travis RC, Campa D, Ingles SA, John EM, Hayes RB, Pharoah PD, Pashayan N, Khaw KT, Stanford JL, Ostrander EA, Signorello LB, Thibodeau SN, Schaid D, Maier C, Vogel W, Kibel AS, Cybulski C, Lubinski J, **Cannon-Albright L**, Brenner H, Park JY, Kaneva R, Batra J, Spurdle AB, Clements JA, Teixeira MR, Dicks E, Lee A, Dunning AM, Baynes C, Conroy D, Maranian MJ, Ahmed S, Govindasami K, Guy M, Wilkinson RA, Sawyer EJ, Morgan A, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As NJ, Woodhouse CJ, Thompson A, Dudderidge T, Ogden C, Cooper CS, Lophatananon A, Cox A, Southey MC, Hopper JL, English DR, Aly M, Adolfsson J, Xu J, Zheng SL, Yeager M, Kaaks R, Diver WR,

Gaudet MM, Stern MC, Corral R, Joshi AD, Shahabi A, Wahlfors T, Tammela TL, Auvinen A, Virtamo J, Klarskov P, Nordestgaard BG, Roder MA, Nielsen SF, Bojesen SE, Siddiq A, Fitzgerald LM, Kolb S, Kwon EM, Karyadi DM, Blot WJ, Zheng W, Cai Q, McDonnell SK, Rinckleb AE, Drake B, Colditz G, Wokolorczyk D, Stephenson RA, Teerlink C, Muller H, Rothenbacher D, Sellers TA, Lin HY, Slavov C, Mitev V, Lose F, Srinivasan S, Maia S, Paulo P, Lange E, Cooney KA, Antoniou AC, Vincent D, Bacot F, Tessier DC, Kote-Jarai Z, Easton DF. (2013). Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet*, *45*(4), 385-91, 391e1-2.

Kote-Jarai Z, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Dadaev T, Jugurnauth-Little S, Ross-Adams H, Al Olama AA, Benlloch S, Halim S, Russel R, Dunning AM, Luccarini C, Dennis J, Neal DE, Hamdy FC, Donovan JL, Muir K, Giles GG, Severi G, Wiklund F, Gronberg H, Haiman CA, Schumacher F, Henderson BE, Le Marchand L, Lindstrom S, Kraft P, Hunter DJ, Gapstur S, Chanock S, Berndt SI, Albanes D, Andriole G, Schleutker J, Weischer M, Canzian F, Riboli E, Key TJ, Travis RC, Campa D, Ingles SA, John EM, Hayes RB, Pharoah P, Khaw KT, Stanford JL, Ostrander EA, Signorello LB, Thibodeau SN, Schaid D, Maier C, Vogel W, Kibel AS, Cybulski C, Lubinski J, **Cannon-Albright L**, Brenner H, Park JY, Kaneva R, Batra J, Spurdle A, Clements JA, Teixeira MR, Govindasami K, Guy M, Wilkinson RA, Sawyer EJ, Morgan A, Dicks E, Baynes C, Conroy D, Bojesen SE, Kaaks R, Vincent D, Bacot F, Tessier DC, Easton DF, Eeles RA. (2013). Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression. *Hum Mol Genet*, *22*(12), 2520-8.

Nelson Q, Agarwal N, Stephenson R, **Cannon-Albright LA**. (2013). A population-based analysis of clustering identifies a strong genetic contribution to lethal prostate cancer. *Front Genet*, *4*, 152

Teerlink C, Farnham J, Allen-Brady K, Camp NJ, Thomas A, Leachman S, **Cannon-Albright L**. (2011). A unique genome-wide association analysis in extended Utah high-risk pedigrees identifies a novel melanoma risk variant on chromosome arm 10q. *Hum Genet*. (Epub ahead of print).

Xu J, Lange EM, Lu L, Zheng SL, Wang Z, Thibodeau SN, **Cannon-Albright LA**, Teerlink CC, Camp NJ, Johnson AM, Zuhlke KA, Stanford JL, Ostrander EA, Wiley KE, Isaacs SD, Walsh PC, Maier C, Luedeke M, Vogel W, Schleutker J, Wahlfors T, Tammela T, Schaid D, McDonnell SK, DeRycke MS, Cancel-Tassin G, Cussenot O, Wiklund F, Grönberg H, Eeles R, Easton D, Kote-Jarai Z, Whittemore AS, Hsieh CL, Giles GG, Hopper JL, Severi G, Catalona WJ, Mandal D, Ledet E, Foulkes WD, Hamel N, Mahle L, Moller P, Powell I, Bailey-Wilson JE, Carpten JD, Seminara D, Cooney KA, Isaacs WB, International Consortium for Prostate Cancer Genetics. (2013). HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet*, *132*(1), 5-14.

Appendix

Data Collection for prostate characteristics phenotype.

DOD project ID number/ MRN:
Age:
Race:
- Caucasian =1
- African American =2
- Hispanic =3
- Asian/Pacific Islander=4
- Southeast Asian=5
- Other=6

Family history of Prostate Cancer:
- No=1
- First degree relative =2
- Other =3

BMI:
- <18.5 (underweight)=1
- 18.5-25 (healthy weight)=2
- 25-30(overweight)=3
- 30-34.99 (obese class 1)=4
- 35-39.99 (obese class 2)=5
- >40 (obese class 3)=6 * WHO classification of BMI

Number of comorbidities:
- None=1
- One=2
- Two=3
- Three or more= 4

History of Smoking:
- Yes= 1
- No= 2

Current Smoking: **2**
- Yes=1
- No=2

Smoking in pack years:
- 1-10=1
- 11-20=2
- 21-30=3
- 31=40=4
- 41-50=5
- >51=6

PSA before diagnosis of Prostate Cancer:

PSA doubling time before diagnosis of Prostate Cancer:

Primary Gleason (needle) Grade:
- o 5=5
- o 4=4
- o 3=3
- o 2= 2
- o 1=1

Secondary Gleason Grade (needle) :
- o 5=5
- o 4=4
- o 3=3
- o 2=2
- o 1=1

Tertiary Gleason Grade (needle):
- o 5=5
- o 4=4
- o 3=3
- o 2=2
- o 1=1

Clinical Stage:
- o T0=1
- o T1a=2
- o T1b=3
- o T1c=4
- o T2=5
- o T2a=6
- o T2b=7
- o T2c=8
- o T3=9
- o T3a=10
- o T3b=11
- o T4=12
- o TX=13

Number of cores:

Actual number of cores involved:

LVI:
- o Absent=1
- o Present = 2

Prostate MRI prior to treatment:
- Yes=1
- No=**2**

MRI positive ECE:
- Positive=1
- Negative =2

Pathological Stage (surgical):
- pT2= 1
- pT2a=2
- pT2b=3
- pT2c=4
- pT3=5
- pT3a=6
- pT3b=7
- pT4=8

Pathological Gleason score (surgical):
- 5=5
- 4=4
- 3=3
- 2=2
- 1=1

Pathological tertiary Gleason score (surgical):
- 5=5
- 4=4
- 3=3
- 2=2
- 1=1

Seminal Vesicle Involvement:
- SVI Type I=1
- SVI Type II=2
- SVI Type III=3 * from *Anatomic considerations in prostate carcinoma,* Department of Urology, Baylor School of Medicine, 1989

Lymph node involvement:
- NX=1
- NO=2
- N1=3

Number of Lymph Nodes involved:

Metastatic disease present at diagnosis:
- Present=1
- Absent=2

PSA doubling time before starting ADT:
- o <3 months= 1
- o 3-6 months = 2
- o 6-9 months = 3
- o 9-12 months= 4
- o >12 months= 5

Time of initiation of androgen deprivation therapy (castration) after definitive therapy:
- o 0-6 months=1
- o 6-12 month=2
- o 12-24 months =3
- o 24 months = 4

PSA after 7 months of castration:
- o <. 02= 1
- o 0.02-4=2
- o 4-10=3
- o 10=4

PSA after 1 year of castration:

Best PSA response to castration:
- o 90%= 1
- o >50%=2
- o >30 %=3
- o No response = 4

Best imaging response to castration:
- o Stable= 1
- o Partial= 2
- o Complete = 3
- o No response=4

Time to PSA progression on castration (25% increase from nadir):
- o 0-3 months=1
- o 3-6 months=2
- o 6-12 months=3
- o 12-18 months=4
- o 18-24 months=5
- o > 24 month=6

Time to imaging or clinical progression on castration:
- o 0-3 months=1
- o 3-6 months=2
- o 6-9 months =3
- o 9-12months =4
- o 12-18 months=5
- o 18-24 months=6
- o > 24 month=7

Line of treatment after onset of castration refractory disease:
- First line=1
- Second line=2
- Third line= 3
- Fourth line=4
- Fifth line=5
- Sixth line=6

Best PSA response to:
- 90%= 1
- >50%=2
- >30 %=3
- No response = 4

Best imaging response to:
- Stable= 1
- Partial= 2
- Complete = 3
- No response =4

Time to PSA progression on drug:
- 0-3 months=1
- 3-6 months=2
- 6-12 months=3
- 12-18 months=4
- 18-24 months=5
- >24 months=6

Time to imaging or clinical progression on:
- 0-3 months=1
- 3-6 months=2
- 6-9 months =3
- 9-12months =4
- 12-18 months=5
- 18-24 months=6
- > 24 month=7

Line of treatment after onset of castration refractory disease::
- First line=1
- Second line=2
- Third line= 3
- Fourth line=4
- Fifth line=5
- Sixth line=6

Best PSA response to:
- > 90%= 1
- >50%=2
- >30 %=3
- No response = 4

Best imaging response to:
- o Stable= 1
- o Partial= 2
- o Complete = 3
- o No response= 4

Time to PSA progression on:
- o 0-3 months=1
- o 3-6 months=2
- o 6-12 months=3
- o 12-18 months=4
- o 18-24 months=5
- o > 24 months=6

Time to imaging or clinical progression on:
- o 0-3 months=1
- o 3-6 months=2
- o 6-9 months =3
- o 9-12months =4
- o 12-18 months=5
- o 18-24 months=6
- o > 24 months=7

Line of treatment after onset of castration refractory disease::
- o First line=1
- o Second line=2
- o Third line= 3
- o Fourth line=4
- o Fifth line=5
- o Sixth line=6

Best PSA response to:
- o > 90%= 1
- o >50%=2
- o >30 %=3
- o No response = 4

Best imaging response to:
- o Stable= 1
- o Partial= 2
- o Complete = 3
- o No response =4

Time to PSA progression on:
- o 0-3 months=1
- o 3-6 months=2
- o 6-12 months=3
- o 12-18 months=4
- o 18-24 months=5
- o > 24 months=6

Time to imaging or clinical progression on:
- o 0-3 months=1
- o 3-6 months=2
- o 6-9 months =3
- o 9-12months =4
- o 12-18 months=5
- o 18-24 months=6
- o > 24 months=7

Line of treatment after onset of castration refractory disease::
- o First line=1
- o Second line=2
- o Third line= 3
- o Fourth line=4
- o Fifth line=5
- o Sixth line=6

Best PSA response to:
- o > 90%= 1
- o >50%=2
- o >30 %=3
- o No response = 4

Best imaging response to:
- o Stable= 1
- o Partial= 2
- o Complete = 3
- o No response =4

Time to PSA progression on:
- o 0-3 months=1
- o 3-6 months=2
- o 6-12 months=3
- o 12-18 months=4
- o 18-24 months=5
- o > 24 months=6

Time to imaging or clinical progression on drug:
- o 0-3 months=1
- o 3-6 months=2
- o 6-9 months =3
- o 9-12months =4
- o 12-18 months=5
- o 18-24 months=6
- o > 24 months=7

Line of treatment after onset of castration refractory disease:
- o First line=1
- o Second line=2
- o Third line= 3
- o Fourth line=4
- o Fifth line=5
- o Sixth line=6

Best PSA response to:
- o > 90%= 1
- o >50%=2
- o >30 %=3
- o No response = 4

Best imaging response to:
- o Stable= 1
- o Partial= 2
- o Complete = 3
- o No response=4

Time to PSA progression on:
- o 0-3 months=1
- o 3-6 months=2
- o 6-12 months=3
- o 12-18 months=4
- o 18-24 months=5
- o > 24 months=6

Time to imaging or clinical progression on:
- o 0-3 months=1
- o 3-6 months=2
- o 6-9 months =3
- o 9-12months =4
- o 12-18 months=5
- o 18-24 months=6
- o 24 months=7

Line of treatment after onset of castration refractory disease:
- o First line=1
- o Second line=2
- o Third line= 3
- o Fourth line=4
- o Fifth line=5
- o Sixth line=6

Best PSA response to:
- o > 90%= 1
- o >50%=2
- o >30 %=3
- o No response = 4

Best imaging response to:
- o Stable= 1
- o Partial= 2
- o Complete = 3
- o No response=4

Time to PSA progression on:
- o 0-3 months=1
- o 3-6 months=2
- o 6-12 months=3
- o 12-18 months=4
- o 18-24 months=5
- o > 24 months=6

Time to imaging or clinical progression on:
- o 0-3 months=1
- o 3-6 months=2
- o 6-9 months =3
- o 9-12months =4
- o 12-18 months=5
- o 18-24 months=6
- o > 24 months=7

Concomitant bisphosphonates:
- o Yes=1
- o No=2

Concomitant RANKL inhibitor:
- o Yes =1
- o No=2

Pathologic fractures:
- o Yes=1
- o No=2

Bone Density:
- o Normal=1
- o Osteopenia=2
- o Osteoporosis=3

Overall Survival:
- o # of months